Homework15

Soomin Park

Read the data file

```
data <- read.table("results.tsv")
head(data)
```

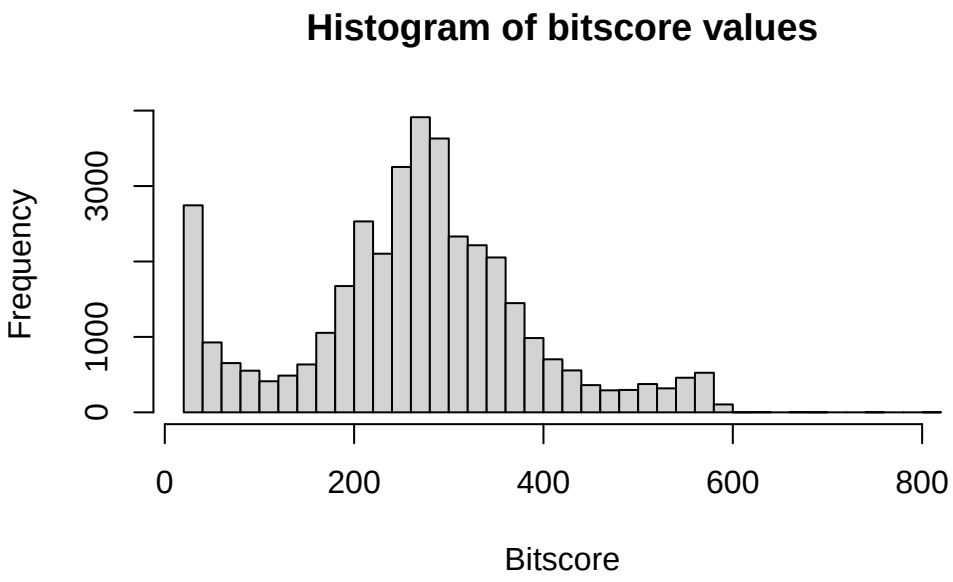
```
          V1             V2     V3  V4  V5 V6 V7  V8  V9 V10      V11   V12
1 NP_598866.1 XP_009294521.1 46.154 273 130  6  4 267 420 684 1.69e-63 214.0
2 NP_598866.1 NP_001313634.1 46.154 273 130  6  4 267 476 740 4.51e-63 214.0
3 NP_598866.1 XP_009294513.1 46.154 273 130  6  4 267 475 739 4.69e-63 214.0
4 NP_598866.1 NP_001186666.1 33.071 127  76  5  4 126 338 459 5.19e-12  67.8
5 NP_598866.1 NP_001003517.1 30.400 125  82  4  4 126 344 465 2.66e-11  65.5
6 NP_598866.1 NP_001003517.1 30.645  62  41  2 53 113  43 103 4.40e-01  33.9
```

Set column names

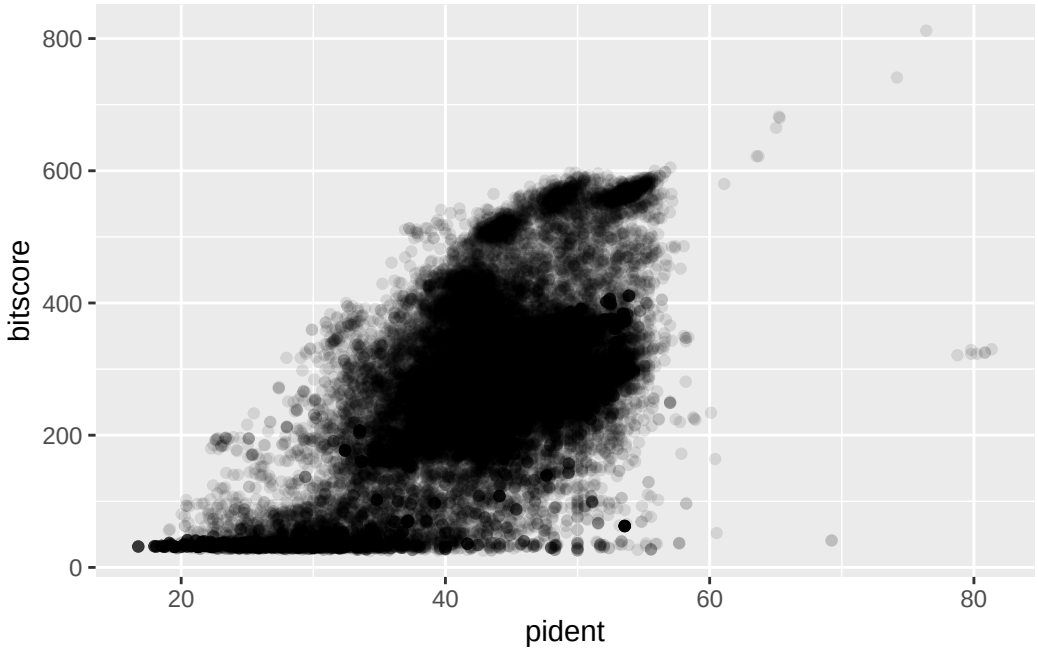
```
colnames(data) <- c("qseqid", "sseqid", "pident", "length", "mismatch", "gapopen", "qstart
```

Generate histogram

```
hist(data$bitscore, breaks = 30, main = "Histogram of bitscore values",
     xlab = "Bitscore", ylab = "Frequency")
```




```
library(ggplot2)
ggplot(data, aes(pident, bitscore)) + geom_point(alpha=0.1)
```




```
ggplot(data, aes((data$pident * (data$qend - data$qstart)), bitscore)) + geom_point(alpha=
```

```
Warning: Use of `data$pident` is discouraged.
i Use `pident` instead.

Warning: Use of `data$qend` is discouraged.
i Use `qend` instead.

Warning: Use of `data$qstart` is discouraged.
i Use `qstart` instead.

Warning: Use of `data$pident` is discouraged.
i Use `pident` instead.

Warning: Use of `data$qend` is discouraged.
i Use `qend` instead.

Warning: Use of `data$qstart` is discouraged.
i Use `qstart` instead.
```

```
`geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```